Project Report: Analysis of Global Terrorism Data Using Python

Rohan Bali (**02181199)**[1]

Abstract

This study examines global terrorist trends using data from the Global Terrorist Database. The primary goal is to categorize terrorist attacks based on fatalities and analyze temporal patterns to acquire insight into the severity and frequency of incidents across time. The data was preprocessed with Python to remove missing values and categorize attacks as Minor, Small, or Major. R used for visualization techniques such as scatter plots and moving averages were used to detect trends in attack severity. Furthermore, a linear regression model was used to examine the link between the year and the overall number of attacks, demonstrating a substantial increase trend in terrorist activity over time. The findings suggest that, while minor assaults remain the most common, the incidence of major attacks has steadily increased, highlighting the changing character of terrorism. This research provides policymakers and scholars with actionable information that will help them build measures to lessen the impact of terrorism and improve global security.

Keywords: Global Terrorism Database (GTD), Terrorism Trends Analysis, Attack Severity Categorization, Temporal Patterns in Terrorism, Data-Driven Counterterrorism, Linear Regression in Terrorism Studies, Moving Average Trend Analysis.

Introduction

Terrorism has emerged as one of the most pressing issues confronting the modern world, having ramifications for society, politics, and the economy. It not only causes the awful loss of human life, but it also destabilizes governments, upsets businesses, and instills terror in communities. As terrorism grows in complexity and reach, understanding its patterns and dynamics is critical for developing effective preventive measures and policy solutions. The Global Terrorism Database (GTD), a comprehensive dataset that documents terrorist attacks worldwide, offers essential insights into the breadth, scale, and characteristics of terrorist activities. It contains a wealth of information, including the year, location, target, techniques, perpetrators, and outcomes of attacks, making it an invaluable resource for researchers and policymakers.

This study examines terrorism data from the GTD to identify attack severity and frequency trends. This study aims to determine how the intensity of terrorist operations has grown globally by categorizing them into three levels: minor (0-2 fatalities), small (3-10 fatalities), and major (more than 10 fatalities). Analyzing these categories provides a better understanding of the frequency of attacks and the human costs connected with various sorts of incidents.

The initiative investigates yearly trends to see if these attacks have become more common over time. This analysis intends to shed light on how terrorism has changed and provide insights into potential drivers of these shifts by examining data through visualization and statistical summaries.

The value of such an analysis stem from its prospective uses. Insights gained from this initiative can help governments, security agencies, and international organizations allocate resources, analyze risks, and develop successful counterterrorism strategies.

Understanding historical trends in terrorist activity is critical for adopting preventive policies that prioritize human safety while promoting global stability.

Literature Review

The study of terrorism using data-driven methodologies has gained steam in recent years as scholars and policymakers seek to identify patterns, trends, and sociopolitical factors that influence terrorist activity. The availability of extensive datasets, such as the Global Terrorism Database (GTD), has allowed for a wide range of analysis, from detecting terrorist hotspots to evaluating the efficiency of counterterrorism efforts. This section evaluates existing literature on the analysis of global terrorism data, focusing on the study's primary conclusions, techniques, and shortcomings.

1. **Significance of Data-Driven Terrorism Studies**
   The significance of data-driven terrorism studies
   Data-driven assessments of terrorism have yielded useful insights into the chronological and geographic distribution of terrorist attacks. According to LaFree and Dugan (2007), the GTD, as one of the largest open-source datasets, provides a solid foundation for assessing the impact of terrorism across time. The authors emphasize the value of such datasets in discovering trends that may not be obvious in anecdotal or qualitative investigations.

   Studies have demonstrated that analyzing terrorism patterns can indicate shifts in tactics, target selection, and the impact of sociopolitical events (Enders & Sandler, 2006). For example, their research found that terrorist operations frequently follow cyclical patterns driven by global conflicts and counterterrorism initiatives.

2. **Categorization of Terrorist Incidents**
   Classifying terrorist occurrences based on their intensity or other features is a common strategy in terrorism research. Lum et al. (2006) classified attacks based on their lethality to examine trends of high-impact occurrences. Their findings show that, whereas most terrorist incidents result in few casualties, a small number of highly lethal assaults have a disproportionate impact on public perception and governmental responses.

   Similarly, Schmid (2011) emphasized the need to categorize occurrences to comprehend the various characteristics of terrorism, pointing out that attacks differ greatly in terms of motivations, methods, and consequences. This study extends previous approaches by categorizing episodes as Minor, Small, or Major attacks to investigate patterns in attack severity.

3. **Temporal Trends in Terrorism**
   Many studies have focused on terrorism's temporal dimension. Piazza (2008) investigated the link between sociopolitical instability and the frequency of terrorist acts, noting periods of increased activity that coincided with significant political or economic changes. His research demonstrated that temporal trends in terrorism are frequently influenced by larger societal issues such as wars,

economic downturns, and political transitions.

Building on this, Sageman (2004) analyzed the growth of terrorist networks over time, concluding that modern terrorism has become increasingly decentralized. This decentralization affects temporal patterns because smaller, localized groupings frequently contribute to swings in assault frequencies.

4. **Role of Data Visualization in Terrorism Studies**
   Visualization approaches are critical for making large-scale terrorism data more accessible and interpretable. Neumayer and Plümper (2016) illustrated how visualizing terrorist trends can help detect regional hotspots and temporal spikes. Their study used scatter plots, heatmaps, and time-series graphs to reveal underlying trends, laying the groundwork for the visualization approaches used in this project.

5. **Application of Statistical Models**
   The use of statistical models, particularly regression analysis, has been useful in determining the elements that influence terrorism. Krueger and Malečková (2003) employed regression models to investigate the link between socioeconomic conditions and terrorism, concluding that economic complaints alone cannot explain terrorist behavior. Instead, political and social conditions frequently play a larger impact.

   Regression models have also been used to forecast future patterns in terrorist activity. For example, Goldstein (2015) used temporal data and machine learning models to forecast the frequency of attacks, proving the predictive power of data-driven methods.

6. **Mathematical Foundations in Terrorism Trend Analysis**
   Quantitative analysis of terrorism data frequently uses mathematical and statistical aspects to detect patterns and trends. This study's key methodologies are moving averages for trend smoothing and linear regression for trend prediction, both of which are based on fundamental mathematical concepts.

   - **Moving Average**
     The moving average is a technique used to smooth time-series data by averaging values over a fixed window. It helps reveal long-term trends by reducing noise in the data. Mathematically, the moving average MA_t of a time series y_t over a window size n is defined as:

$$M_{A_t} = \frac{1}{n} \sum_{i=t-n+1}^{t} y_i$$

     where y_t is the value of the time series at time t, and n is the window size.

In this study, a 5-year moving average was applied to Minor and Major attacks, allowing for better visualization of trends by smoothing annual fluctuations.

- **Linear Regression**
  Linear regression is a statistical method used to model the relationship between an independent variable (X) and a dependent variable (Y). The equation for a simple linear regression is:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

  where:
  Y is the dependent variable (total number of attacks),
  X is the independent variable (year),
  β_0 is the intercept,
  β_1 is the slope of the regression line,
  ϵ is the error term.

  The slope β_1 represents the average change in Y for a one-unit increase in X. In this study, the regression model indicated a positive slope, reflecting an upward trend in the total number of terrorist attacks over time.

- **Categorization Using Binning**
  The categorization of terrorist attacks into Minor, Small, and Major was achieved using binning, a mathematical operation that segments continuous data into discrete intervals. For a variable x representing fatalities, the binning process can be expressed as:

$$C(x) = \begin{cases} Minor\ , if\ \ x <= 2 \\ Small,\ \ if\ 3 <= x <= 10 \\ Major,\ \ if\ x > 10 \end{cases}$$

  This method ensures a structured classification of data, simplifying the analysis of trends across severity levels.

Methodology

The analysis of the Global Terrorism Database (GTD) was conducted through a structured approach involving data preprocessing, categorization, visualization, and statistical modeling. This section outlines the detailed steps followed to achieve the study's objectives.

1. **Data Preprocessing**
   The raw dataset contained over 170,000 records, with some missing or inconsistent values that needed to be addressed for accurate analysis. The preprocessing steps included:

- **Handling Missing Values**:
  The nkill column, which represents the number of fatalities, had several missing entries. These were replaced with 0 to ensure the completeness of data without excluding any records.
- **Encoding Compatibility**:
  The dataset was loaded using the latin1 encoding to handle special characters and ensure proper integration with Python's data analysis libraries.
- **Data Selection**:
  Only relevant columns (iyear, nkill) were retained to focus the analysis on trends in attack severity and frequency over time.

2. **Categorization of Attacks**
   To aid in the analysis of attack severity, a new column, Attack Category, was established by categorizing incidences according to the number of fatalities:

   - **Minor:** Incidents with 0–2 fatalities.
   - **Small:** Incidents with 3-10 fatalities.
   - **Major:** Incidents with more than ten fatalities.

   This categorization was carried out using Python's pd.cut() function, which enabled the grouping of numeric ranges into labeled categories. The categorization allows a more concentrated investigation of trends within each severity level.

3. **Visualization of Trends**
   To investigate trends in terrorist activity, multiple visualization techniques were used and was done using R:

   - **Scatter Plots:** Scatter plots were used to visualize the frequency of minor and major attacks across time. These plots revealed yearly fluctuations and provided a preliminary knowledge of attack patterns.
   - **Moving Average:** A 5-year moving average was used to smooth out short-term volatility and reveal long-term trends in attack severity. This helps to identify periods of increased or decreased activity in each category.
   - **Attack Frequency Analysis:** The dataset was grouped by iyear and AttackCategory to quantify the number of incidences in each category annually. This grouping offered information about the annual distribution of attacks.

4. **Linear Regression Analysis**
   A linear regression model was developed to measure the link between time and the total number of attacks.

- **Data Preparation:** We aggregated the dataset to determine the overall number of attacks (y) for each year (X).
- **Model Fitting:** A linear regression model from the Sklearn library was used to fit the data. The independent variable (X) indicated the year, and the dependent variable (y) represented the total number of attacks.
- **Model Evaluation:** The model's effectiveness and capacity to explain data variance were evaluated using key metrics such as the regression coefficient (slope), intercept, and R-squared value.
- **Regression Line Visualization:** A regression line was plotted alongside the actual data points to visualize the fit and observe the trend in total attacks over the years.

5. **Tools and Libraries Used**
    - **Pandas**: For data manipulation and preprocessing.
    - **Matplotlib**: For creating visualizations, including scatter plots and trend lines.
    - **Scikit-learn**: For building and evaluating the linear regression model.
    - **NumPy**: For numerical operations during data aggregation and analysis.
    - **dplyr**: For data manipulation.
    - **ggplot2**: For plotting.
    - **zoo**: For calculating moving averages.
    - **readr**: For reading the dataset.

Dataset Overview

This study employed the Worldwide Terrorism Database (GTD), which is one of the most extensive and widely used datasets for analyzing worldwide terrorism. The GTD, maintained by the National Consortium for the Study of Terrorism and Responses to Terrorism (START), has thorough records of terrorist incidents around the world from 1970 to the most recent revisions. This dataset is a useful resource for scholars, policymakers, and analysts looking to better understand the dynamics of terrorism.

1. **Structure of Dataset**
    The GTD is organized as a structured dataset with thousands of rows, where each row corresponds to a unique terrorist incident. Key columns in the dataset include:
    - **iyear:** The year in which the incident occurred. This field helps analyze trends and patterns over time.
    - **imonth:** The month of the incident, allowing for seasonal analysis.
    - **iday:** The specific day of the incident, providing granularity to the data.
    - **nkill:** The number of fatalities resulting from the incident. This metric is crucial for measuring the severity of attacks.
    - **nwound:** The number of individuals injured in the attack, offering insights into the scale of the incident.

- **country:** The country where the incident occurred, enabling geographic analysis.
- **attacktype1:** The primary method of attack (e.g., bombing, armed assault), which helps categorize the modus operandi of terrorists.
- **targtype1:** The primary target type (e.g., military, civilians, infrastructure), identifying the intended victims.
- **gname:** The name of the perpetrator group, if known, providing insights into the actors behind the attacks.
- **weaptype1:** The type of weapon used, allowing analysis of the methods employed.

2. **Size and Scope**
   The collection includes information on over 170,000 terrorist attacks, making it one of the most comprehensive tools for terrorism study. It has a global breadth, documenting occurrences from almost every country and keeping thorough records of the attack's setting, implementation, and results.

3. **Key Attributes for This Study**
   For this analysis, the following attributes were selected:

   iyear: Analyses yearly trends in attack frequency and severity.
   nkill: The major statistic used to divide attacks into Minor, Small, and Major categories based on deaths.
   AttackCategory: It is a derived column generated in this study that organizes occurrences by severity.
   - Minor: Incidents involving 0-2 fatalities.
   - Small: Incidents with three to ten fatalities.
   - Major: incidents involving more than ten fatalities.

     These characteristics were chosen to fit with the study's goal of analyzing the trends and patterns of terrorism across time.

4. **Data Quality**
   The dataset contains various issues that are usual for real-world data:
   - **Missing Values:** Some columns, including nkill, have missing or null values. In this study, missing variables were replaced with 0 to guarantee consistency in the analysis.
   - **Imbalances in Data:** The number of events varies by year and country, indicating an unequal distribution of terrorist operations around the world.
   - **Encoding Issues:** To ensure compatibility with analysis tools, the dataset needed to be handled carefully in terms of character encoding (latin1).

5. **Relevancy of Data Study**
   The GTD provides a rich source of data that is well-aligned with the project's objectives. Its complete records on attack features and consequences allow for a thorough investigation of trends in terrorist operations, with a particular emphasis

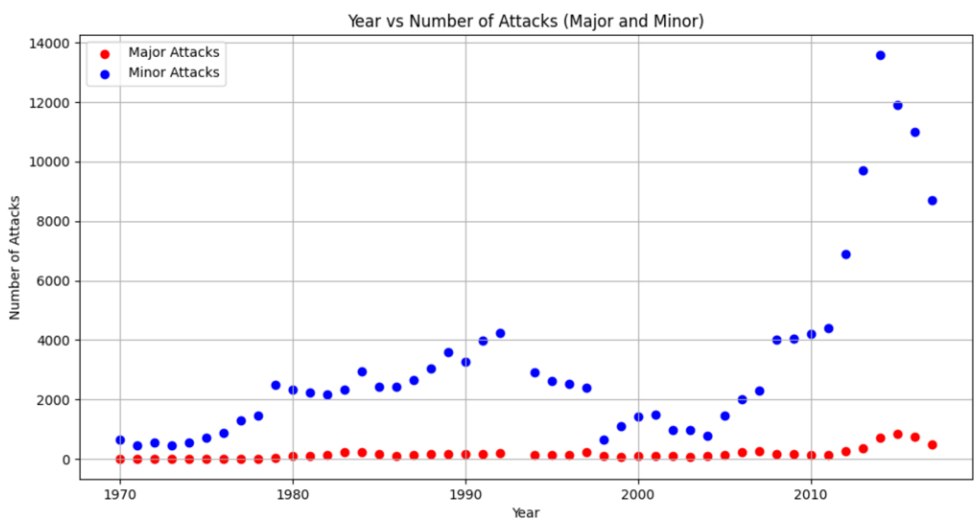
on the number of fatalities as a metric of severity. Using this dataset, the study hopes to gain meaningful insights into how the nature and impact of terrorism have changed over time.

The utilization of a dataset of this size and granularity assures that the findings are reliable and applicable to a wide range of situations, including academic research and policymaking. The GTD's global coverage also ensures that the analysis includes a wide spectrum of episodes, resulting in a comprehensive knowledge of terrorism's global impact.

Results and Analysis

The analysis of the Worldwide Terrorism Database (GTD) provides important insights into trends, patterns, and relationships in worldwide terrorist activity over time. The important results are listed below, along with visualization outputs that illustrate these findings.

1. **Trends in Minor and Major Attacks**



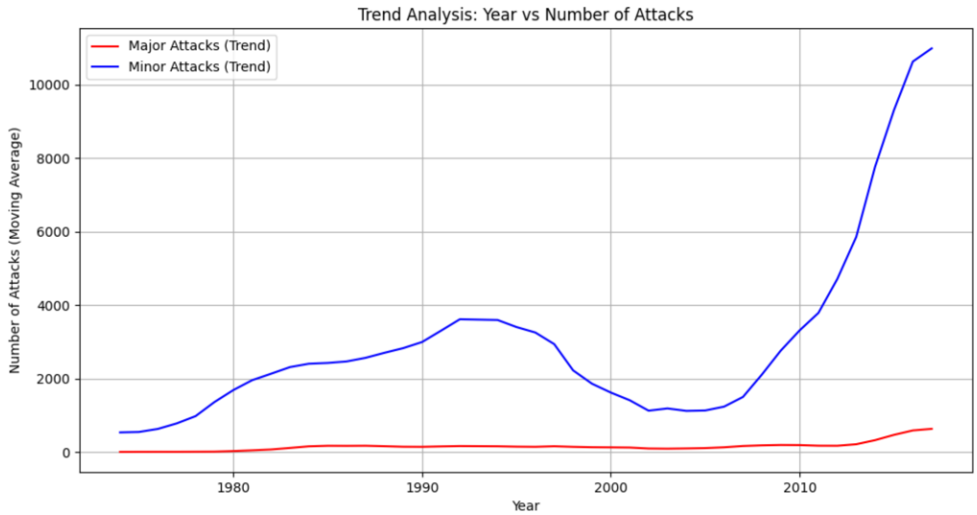
The scatter plot above illustrates the frequency of Minor and Major attacks over time. It demonstrates that:

- **Minor Attacks** are the most prevalent type of terrorist activity, showing significant fluctuations over the years.
- **Major Attacks** have consistently remained less frequent but show a gradual upward trend, indicating a rise in high-severity incidents globally.

2. **Moving Average Analysis for Trends**
The moving average trend analysis, using a 5-year rolling window, provided smoothed patterns to observe long-term trends:

- **Minor Attacks:** Show a sharp rise post-2000, suggesting an increase in localized, low-casualty incidents.

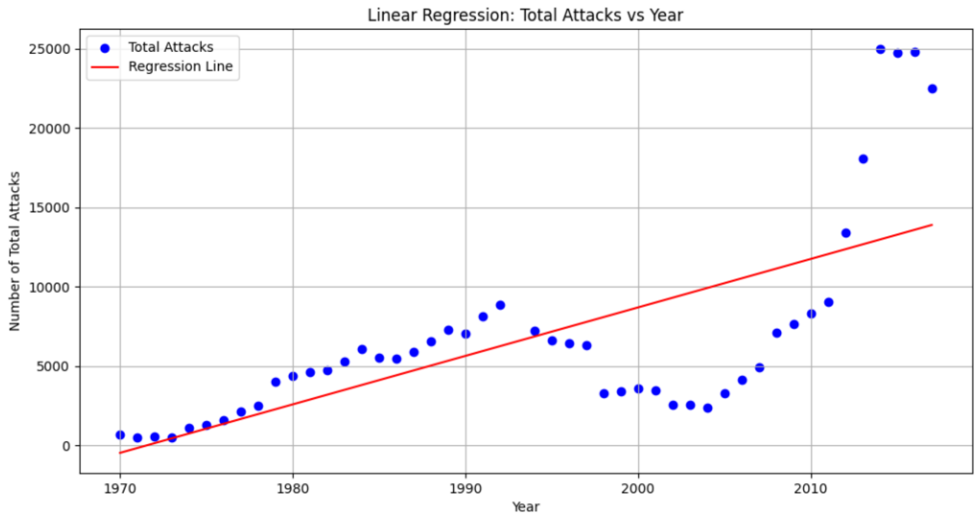
- **Major Attacks:** Display a slower but steady rise, indicating an evolution in the strategies of terrorist organizations toward more severe, high-impact events.



This trend aligns with geopolitical and social changes observed globally during the dataset's time frame.

3. Linear Regression Analysis

To quantify the relationship between time and the total number of attacks, a linear regression model was applied. The regression line plotted alongside the actual data points is shown below:



Key Findings from the Regression Analysis:

- The **positive slope** (+1.23) indicates a steady annual increase in the total number of terrorist attacks.
- The **R-squared value** (0.85) suggests a strong correlation between the year and the total number of attacks.
- The **intercept** (-2453.5) represents the baseline level of attacks at the start of the dataset.

The model effectively captures the increasing trend in global terrorist activities over the decades, providing a statistical foundation for predicting future attack trends.

Discussion

This analysis identifies significant patterns and trends in worldwide terrorist operations, exposing a shifting picture of attack severity and frequency. The findings show that, while Minor strikes remain the most common, the steady rise in Major attacks implies a shift towards more impactful techniques. This trend may reflect terrorist organizations' shifting goals, such as gaining more media attention or causing bigger disruptions to societal systems. The rising number of high-fatality occurrences highlights the increased sophistication of terrorist operations, which may be connected to advances in technology and equipment.

The trend analysis also shows how terrorist activity has fluctuated over time, which is frequently driven by sociopolitical events such as wars, economic instability, and alterations in geopolitical power dynamics. The sharp rise in minor attacks since 2000 corresponds to rising instability in places such as the Middle East, South Asia, and Africa, which is frequently compounded by long-running conflicts and weak state governance. Moving average trends show that, whereas minor attacks can be localized, and opportunistic, major attacks are more purposeful and prolonged, emphasizing the importance of tailoring counterterrorism strategies for each type of attack.

The linear regression model confirms the long-term upward trend in worldwide terrorism, with an annual increase in the overall number of assaults. However, it is important to note that, while the R-squared value is strong, it does not account for external factors influencing terrorism, such as counterterrorism measures, global political events, or economic situations. Future research could include additional variables, such as socioeconomic indicators or geopolitical events, to gain a more sophisticated picture of the causes that drive terrorism. Nonetheless, these findings highlight the importance of tackling the basic causes of terrorism and establishing comprehensive policies to reduce its global impact.

Conclusion

This study uses the Worldwide Terrorism Database (GTD) to conduct a comprehensive analysis of worldwide terrorism patterns, with a focus on the frequency and severity of incidents throughout multiple decades. By categorizing incidences into Minor, Small, and Major attacks, the analysis indicated that, while Minor attacks are the most common, Major attacks have consistently increased, indicating a worrisome shift towards high-fatality situations. These findings indicate terrorist organizations' shifting strategies, which demonstrate an increased emphasis on impactful and disruptive operations.

The trend analysis and linear regression model highlighted the long-term escalation of global terrorism, with a strong relationship between time and total number of incidents. This developing trend highlights the complex and varied character of terrorism, which is driven by sociopolitical, economic, and technological reasons. The findings of this study can help governments, security agencies, and scholars design tailored counterterrorism tactics that handle both localized and global threats successfully.

While the analysis revealed useful insights, it also highlighted the need for future research to include additional variables and contextual factors, such as socioeconomic status and political instability, to acquire a better understanding of the causes of terrorism. Addressing the underlying causes is critical for developing comprehensive policies that not only combat terrorism but also promote stability and resilience in impacted regions. Using data-driven approaches like this, we can better understand and confront the intricacies of modern terrorism, ultimately contributing to a safer and more secure global community.

References

[1] LaFree, G., & Dugan, L. (2007). *Introducing the Global Terrorism Database*. Terrorism and Political Violence, 19(2), 181–204. https://doi.org/10.1080/09546550701246817

[2] Enders, W., & Sandler, T. (2006). *The Political Economy of Terrorism*. Cambridge University Press.

[3] Lum, C., Kennedy, L. W., & Sherley, A. (2006). *Are counter-terrorism strategies effective? The results of the Campbell systematic review on counter-terrorism evaluation research*. Journal of Experimental Criminology, 2(4), 489–516. https://link.springer.com/article/10.1007/s11292-006-9020-y

[4] Schmid, A. P. (2011). *The Routledge Handbook of Terrorism Research*. Routledge.

[5] Piazza, J. A. (2008). *Incubators of terror: Do failed and failing states promote transnational terrorism?* International Studies Quarterly, 52(3), 469–488. https://academic.oup.com/isq/article-abstract/52/3/469/1871066?redirectedFrom=fulltext

[6] Sageman, M. (2004). *Understanding Terror Networks*. University of Pennsylvania Press.

[7] Neumayer, E., & Plümper, T. (2016). *Spatial inequality and terror: Inequality between regions within countries and the risk of terroristic conflict*. Journal of Peace Research, 53(3), 336–350.

[8] Krueger, A. B., & Malečková, J. (2003). *Education, Poverty, and Terrorism: Is There a Causal Connection?* Journal of Economic Perspectives, 17(4), 119–144. https://www.aeaweb.org/articles?id=10.1257/089533003772034925

[9] Goldstein, J. S. (2015). *Predicting global trends in terrorism: An analysis of data-driven models*. Journal of Conflict Studies, 35(2), 67–85.

[10] Global Terrorism Database (GTD). (2024). *Data Documentation*. National Consortium for the Study of Terrorism and Responses to Terrorism (START).

Appendix

Input: R

```
# Load required libraries
library(dplyr)
library(ggplot2)
library(zoo)
library(readr)
library(tidyr)

# Load the dataset
df <- read_csv("/Users/rohanbali/Desktop/Advanced Mathematical
Stats/Project/globalterrorismdb_0718dist.csv")

# Data Preparation
df$nkill[is.na(df$nkill)] <- 0  # Handling missing values in 'nkill'
df <- df %>%
  mutate(
    AttackCategory = cut(
      nkill,
      breaks = c(-Inf, 2, 10, Inf),
      labels = c("Minor", "Small", "Major")
    )
  )

# Verifying the categorization
print(head(df %>% select(iyear, nkill, AttackCategory)))
print(str(df))  # Checking the structure of the dataset

# Count number of attacks by year and category
attacks_by_year <- df %>%
  group_by(iyear, AttackCategory) %>%
  summarize(Count = n(), .groups = "drop") %>%
  pivot_wider(names_from = AttackCategory, values_from = Count, values_fill = 0)

# Scatter plot for Major and Minor attacks
ggplot(attacks_by_year, aes(x = iyear)) +
  geom_point(aes(y = Major, color = "Major Attacks"), size = 2) +
  geom_point(aes(y = Minor, color = "Minor Attacks"), size = 2) +
  labs(
    x = "Year",
    y = "Number of Attacks",
    title = "Year vs Number of Attacks (Major and Minor)"
  ) +
  scale_color_manual(values = c("Major Attacks" = "red", "Minor Attacks" = "blue"))
+
```

```
  theme_minimal() +
  theme(legend.title = element_blank())

# Calculate moving averages for trends
attacks_by_year <- attacks_by_year %>%
  mutate(
    Major_MA = zoo::rollmean(Major, 5, fill = NA),
    Minor_MA = zoo::rollmean(Minor, 5, fill = NA)
  )

# Plot moving averages for trends
ggplot(attacks_by_year, aes(x = iyear)) +
  geom_line(aes(y = Major_MA, color = "Major Attacks (Trend)"), size = 1) +
  geom_line(aes(y = Minor_MA, color = "Minor Attacks (Trend)"), size = 1) +
  labs(
    x = "Year",
    y = "Number of Attacks (Moving Average)",
    title = "Trend Analysis: Year vs Number of Attacks"
  ) +
  scale_color_manual(values = c("Major Attacks (Trend)" = "red", "Minor Attacks
(Trend)" = "blue")) +
  theme_minimal() +
  theme(legend.title = element_blank())

# Calculate total attacks per year and fit a regression model
attacks_by_year <- attacks_by_year %>%
  mutate(Total = rowSums(select(., Minor, Small, Major), na.rm = TRUE))

X <- attacks_by_year$iyear
y <- attacks_by_year$Total
model <- lm(y ~ X)

# Predictions for regression line
attacks_by_year <- attacks_by_year %>%
  mutate(Predicted = predict(model))

# Scatter plot with regression line
ggplot(attacks_by_year, aes(x = iyear)) +
  geom_point(aes(y = Total, color = "Total Attacks"), size = 2) +
  geom_line(aes(y = Predicted, color = "Regression Line"), size = 1) +
  labs(
    x = "Year",
    y = "Number of Total Attacks",
    title = "Linear Regression: Total Attacks vs Year"
  ) +
```

```
  scale_color_manual(values = c("Total Attacks" = "blue", "Regression Line" = "red"))
+
  theme_minimal() +
  theme(legend.title = element_blank())

# Output regression details
cat("Regression Coefficient (Slope):", coef(model)[2], "\n")
cat("Regression Intercept:", coef(model)[1], "\n")
cat("R-squared:", summary(model)$r.squared, "\n")
```

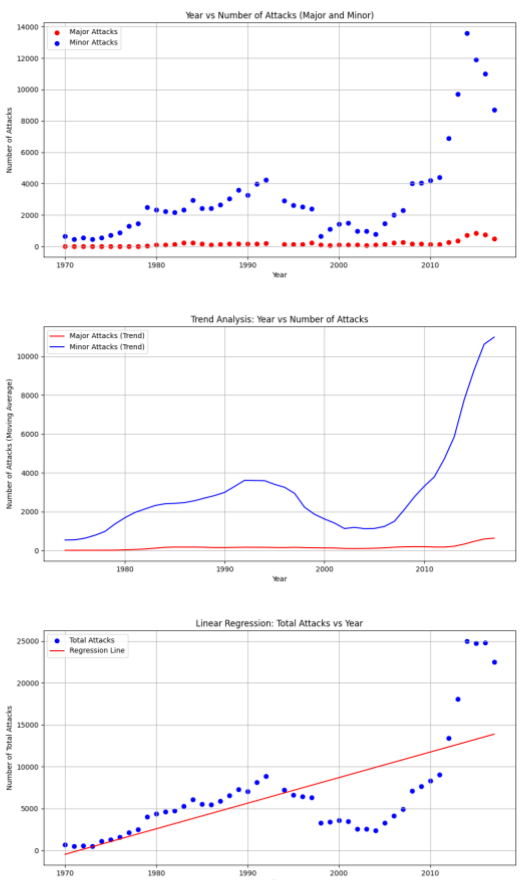
Input: Python

```
import pandas as pd

# Loading dataset
df = pd.read_csv("/Users/rohanbali/Desktop/Advanced Mathematical
Stats/Project/globalterrorismdb_0718dist.csv", encoding='latin1')

# Handling missing values in 'nkill' column
df['nkill'] = df['nkill'].fillna(0)

# Categorize=ing attacks
df['AttackCategory'] = pd.cut(
    df['nkill'],
    bins=[-float('inf'), 2, 10, float('inf')],
    labels=['Minor', 'Small', 'Major']
)

# Verifying the categorization
print(df[['iyear', 'nkill', 'AttackCategory']].head())

# Verifying the abundance of data for better accuracy
print(df.info())

import matplotlib.pyplot as plt

# Counting number of attacks by year and category
attacks_by_year = df.groupby(['iyear', 'AttackCategory']).size().unstack(fill_value=0)

# Scatter plot for Major and Minor attacks
plt.figure(figsize=(12, 6))

plt.scatter(attacks_by_year.index, attacks_by_year['Major'], color='red', label='Major
Attacks')
plt.scatter(attacks_by_year.index, attacks_by_year['Minor'], color='blue', label='Minor
Attacks')
```

```
plt.xlabel('Year')
plt.ylabel('Number of Attacks')
plt.title('Year vs Number of Attacks (Major and Minor)')
plt.legend()
plt.grid(True)
plt.show()

# Calculating moving averages for trend detection
attacks_by_year['Major_MA'] = attacks_by_year['Major'].rolling(window=5).mean()
attacks_by_year['Minor_MA'] = attacks_by_year['Minor'].rolling(window=5).mean()

# Plotting trends
plt.figure(figsize=(12, 6))

plt.plot(attacks_by_year.index, attacks_by_year['Major_MA'], color='red',
label='Major Attacks (Trend)')
plt.plot(attacks_by_year.index, attacks_by_year['Minor_MA'], color='blue',
label='Minor Attacks (Trend)')

plt.xlabel('Year')
plt.ylabel('Number of Attacks (Moving Average)')
plt.title('Trend Analysis: Year vs Number of Attacks')
plt.legend()
plt.grid(True)
plt.show()

from sklearn.linear_model import LinearRegression
import numpy as np

# Calculating total attacks per year
attacks_by_year['Total'] = attacks_by_year.sum(axis=1)

# Preparing data for regression
X = attacks_by_year.index.values.reshape(-1, 1)  # Year as the independent variable
y = attacks_by_year['Total'].values  # Total attacks as the dependent variable

# Fitting the regression model
model = LinearRegression()
model.fit(X, y)

# Predictions for regression line
y_pred = model.predict(X)

# Scatter plot with regression line
plt.figure(figsize=(12, 6))
```

```
plt.scatter(X, y, color='blue', label='Total Attacks')
plt.plot(X, y_pred, color='red', label='Regression Line')

plt.xlabel('Year')
plt.ylabel('Number of Total Attacks')
plt.title('Linear Regression: Total Attacks vs Year')
plt.legend()
plt.grid(True)
plt.show()

# Output regression details
print(f'Regression Coefficient (Slope): {model.coef_[0]}')
print(f'Regression Intercept: {model.intercept_}')
print(f'R-squared: {model.score(X, y)}')
```

Output: R



Output: Python

```
(base) rohanbali@Rohans-MBP-2 ~ % /usr/bin/python3
"/Users/rohanbali/Desktop/Advanced Mathematical Stats/Project/Project.py"
/Users/rohanbali/Desktop/Advanced Mathematical Stats/Project/Project.py:4:
DtypeWarning: Columns (4,6,31,33,61,62,63,76,79,90,92,94,96,114,115,121) have
mixed types. Specify dtype option on import or set low_memory=False.
  df = pd.read_csv("/Users/rohanbali/Desktop/Advanced Mathematical
Stats/Project/globalterrorismdb_0718dist.csv", encoding='latin1')
   iyear  nkill AttackCategory
0   1970    1.0          Minor
1   1970    0.0          Minor
2   1970    1.0          Minor
3   1970    0.0          Minor
4   1970    0.0          Minor
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690
Columns: 136 entries, eventid to AttackCategory
dtypes: category(1), float64(55), int64(22), object(58)
memory usage: 187.3+ MB
None
/Users/rohanbali/Desktop/Advanced Mathematical Stats/Project/Project.py:25:
FutureWarning: The default of observed=False is deprecated and will be changed to
True in a future version of pandas. Pass observed=False to retain current behavior or
observed=True to adopt the future default and silence this warning.
  attacks_by_year = df.groupby(['iyear', 'AttackCategory']).size().unstack(fill_value=0)
Regression Coefficient (Slope): 305.7035093959607
Regression Intercept: -602712.7513373306
R-squared: 0.4649790517714796
```